Seminar Algorithmische Technik
Algorithmic Methods in the Humanities

Introduction and Topic Distribution

Gregor Betz · Michael Hamann · Tamara Mchedlidze
Benjamin Niedermann · Ignaz Rutter

21.4.2016
Overview

1. Organizational issues
   - Structure
   - Requirements

2. Topics
   - Presentations
   - Distribution
Supervisors

Prof. Dr. Gregor Betz
Michael Hamann
Dr. Tamara Mchedlidze
Benjamin Niedermann
Dr. Ignaz Rutter

Institute of Philosophy
Institute of Theoretical Informatics

Participants

Short presentation:
- Name
- Direction of studies/Semester
- Background in Algorithms and Humanities
Learning Goals

- **independent work** on a recent research topic
- extraction of the **highlights** of the topic and their **short** presentation
- investigation the topic and **scientific presentation** of it
- **actively discuss** the topics of the other participants
- present the details of the topics in **your own words** in a written document
- **evaluation** of a scientific text
Learning Goals

- **independent work** on a recent research topic
- extraction of the **highlights** of the topic and their **short** presentation
- investigation the topic and **scientific presentation** of it
- **actively discuss** the topics of the other participants
- present the details of the topics in **your own words** in a written document
- **evaluation** of a scientific text

→ Basic skills of scientific work
→ Preparation for the Master thesis and its presentation
→ Opening of your horizont on the various application of computer science
Structure

today: Topic distribution
today: Topic distribution

Get familiar with the topic
literature review

12.5. Short presentations
today: Topic distribution

Get familiar with the topic
literature review

12.5. Short presentations
19.5. Presentations
2.6. Presentations
9.6. Presentations
16.6. Presentations
23.6. Presentations
30.6. Presentations
structure today: Topic distribution

- Get familiar with the topic literature review

- 12.5. Short presentations
- 19.5. Presentations
- 2.6. Presentations
- 9.6. Presentations
- 16.6. Presentations
- 23.6. Presentations
- 30.6. Presentations

- 15.7. Submission of the written documents (first round)
today: Topic distribution

Get familiar with the topic literature review

12.5. Short presentations
19.5. Presentations
2.6. Presentations
9.6. Presentations
16.6. Presentations
23.6. Presentations
30.6. Presentations

15.7. Submission of the written documents (first round)

Review of two other documents

9.6. Presentations

Choice of the material structuring the document, writing

Selection / structuring lecture content, creating the slides
Structure

today: Topic distribution

Get familiar with the topic literature review

12.5. Short presentations
19.5. Presentations
2.6. Presentations
9.6. Presentations
16.6. Presentations
23.6. Presentations
30.6. Presentations

15.7. Submission of the written documents (first round)

Review of two other documents

Selection/structuring lecture content, Creating the slides

Choice of the material structuring the document, writing

7.8. Submission of the reviews

Revision of your own document

31.08. Final submission of the written document
Structure

today: Topic distribution

Get familiar with the topic literature review

12.5. Short presentations
19.5. Presentations
2.6. Presentations
9.6. Presentations
16.6. Presentations
23.6. Presentations
30.6. Presentations

15.7. Submission of the written documents (first round)

Review of two other documents

19.5. Presentations
2.6. Presentations
9.6. Presentations
16.6. Presentations
23.6. Presentations
30.6. Presentations

7.8. Submission of the reviews

Revision of your own document

31.08. Final submission of the written document

Approximate time

4LP = 120h

Read, do research, understand 40h
Making presentations and practicing 30h
Writing and structuring 30h
Read another work and evaluate 10h
Presentations 10h
Requirements

- independent work
- **Short presentation** for the highlights of the topic
- Presentation of the topic in the **main presentation**
- **Presence** on all the presentations and **participation** in the discussions
- **written document** of the topic, formulated in your own words with concentration on your own questions
- **Review/correction** of two documents of other participants
- Following the deadlines
Requirements

- independent work
- **Short presentation** for the highlights of the topic
- Presentation of the topic in the **main presentation**
- **Presence** on all the presentations and **participation** in the discussions
- **written document** of the topic, formulated in your own words with concentration on your own questions
- **Review/correction** of two documents of other participants
- Following the deadlines

Mark

- Quality of the **main presentation** (content, structure, presentation skills) – 60 %
- Quality of the **final written document** – 40 %
- Not-following the deadlines implies you do not get graded!
Requirements

- independent work
- **Short presentation** for the highlights of the topic
- Presentation of the topic in the **main presentation**
- **Presence** on all the presentations and **participation** in the discussions
- **written document** of the topic, formulated in your own words with concentration on your own questions
- **Review/correction** of two documents of other participants
- Following the deadlines

Mark: short presentation and the first version of the document are not graded

- Quality of the **main presentation** (content, structure, presentation skills) – 60 %
- Quality of the **final written document** – 40 %
- Not-following the deadlines implies you do not get graded!
Familiarization phase

1) first look through the paper, and then read thoroughly
Familiarization phase

1) first look through the paper, and then read thoroughly

2) Make overview of the related work
   - Which works and results are cited? → Related Work
   - Which of these are fundamental?
   - What is the state of the art of the research area of the paper?
     → Article search Google Scholar or DBLP; Access from the university network
Familiarization phase

1) first look through the paper, and then read thoroughly

2) Make overview of the related work
   - Which works and results are cited? → Related Work
   - Which of these are fundamental?
   - What is the state of the art of the research area of the paper?
     → Article search Google Scholar or DBLP; Access from the university network

3) Assess the significance of the paper
   - Who cites this paper?
     → in Google Scholar use the function "cited by"
Familiarization phase

1) first look through the paper, and then read thoroughly

2) Make overview of the related work
   - Which works and results are cited? → Related Work
   - Which of these are fundamental?
   - What is the state of the art of the research area of the paper?
   → Article search Google Scholar or DBLP; Access from the university network

3) Assess the significance of the paper
   - Who cites this paper?
   → in Google Scholar use the function ”cited by”

4) What should you read in the literature?
   - Title and Abstract – Is the content relevant?
   - if yes – Introduction, Conclusion, Main results
   - Only if details are relevant – read all
   - Keep notes!
Short Presentation

Content

- „Advertisement“ of the main presentation
- **Motivation**: applications in the humanities that use these techniques
- **What results the paper contains:** Model, Algorithms and used techniques, evaluation, experimentations . . .
Short Presentation

Content

- „Advertisement“ of the main presentation
- **Motivation:** applications in the humanities that use these techniques
- **What results the paper contains:** Model, Algorithms and used techniques, evaluation, experimentations . . .

Form

- 5 Minutes
- structured and clear slides: examples instead of a lot of text, intuition instead of formal definitions
- Use any software you prefer (Ipe, PowerPoint, Keynote)
  *ipe7.sourceforge.net
Main Presentation

**Timing:** 30 minutes + 15 minutes discussion
Main Presentation

**Timing:** 30 minutes + 15 minutes discussion

**Goal:**
- Inform the listened about the details of your topic
- Motivation of the topic
- Keep the listeners “awakened”, arouse their curiosity
Main Presentation

**Timing:** 30 minutes + 15 minutes discussion

**Goal:**
- Inform the listened about the details of your topic
- Motivation of the topic
- Keep the listeners “awakened”, arouse their curiosity

**Struct.:**
- What can be explained clearly in 30 minutes? Make a selection of the essentials issues. **Talk to your class-mate**
- What is your target group?
- Clear, logical structure, concise but illustrative examples
Main Presentation

**Timing:** 30 minutes + 15 minutes discussion

**Goal:**
- Inform the listened about the details of your topic
- Motivation of the topic
- Keep the listeners “awakened”, arouse their curiosity

**Struct.:**
- What can be explained clearly in 30 minutes? Make a selection of the essentials issues. **Talk to your class-mate**
  - What is your target group?
  - Clear, logical structure, concise but illustrative examples

**Slides:**
- bullet points, no complete sentences
- graphics use
- not too many and not too overloaded slides (about 2 min / slide)
- clear design (suitable colors, uniform font, . . . )
Main Presentation

**Timing:** 30 minutes + 15 minutes discussion

**Goal:**
- Inform the listened about the details of your topic
- Motivation of the topic
- Keep the listeners “awakened”, arouse their curiosity

**Struct.:**
- What can be explained clearly in 30 minutes? Make a selection of the essentials issues. **Talk to your class-mate**
- What is your target group?
- Clear, logical structure, concise but illustrative examples

**Slides:**
- bullet points, no complete sentences
- graphics use
- not too many and not too overloaded slides (about 2 min / slide)
- clear design (suitable colors, uniform font, . . .)

**Present.:**
- practice (many times), measure time
- eye contact with the audience
- speak freely, slowly and clearly
- remain calm, control nervousness
Size: 12–15 in a given \LaTeX-format
Write Document

**Size:** 12–15 in a given \LaTeX-format

**Structure:**
- short and clear Abstract
- Introduction, state of art, applications
- Selected topics in detail
- Summary / Conclusion
- complete references (BibTeX)
Written Document

Size: 12–15 in a given \textsc{latex}-format

Structure:
- short and clear Abstract
- Introduction, state of art, applications
- Selected topics in detail
- Summary / Conclusion
- complete references (BibTeX)

Writing:
- Do not copy text: express in your own words
- Logical structure, keep the red line
- Avoid very long sentences
- clear and consistent formulation
- avoid too long paragraphs
- Use pictures
- Cite and specify all sources correctly
- Check grammar and spelling
Mutual Reviews

**Goal:**
- critical reading of scientific texts
- deeper understanding of two other seminar topics give
- you and your class-mates get meaningful feedback and suggestions for improvement
Mutual Reviews

Goal:
- critical reading of scientific texts
- deeper understanding of two other seminar topics give
- you and your class-mates get meaningful feedback and
  suggestions for improvement

Form:
- written statement (form provided)
- short summary of the content
- strengths and weaknesses of the work
- review of the text (comprehensibility, structure, accuracy,
  language, topic coverage, ambiguities, . . . )
- detailed comments and correction instructions
- as detailed as you would like to get review for your work
- anonymous, objective and fair
Supervision

Your supervisor is your contact in all matters, both regarding the content, topics and the preparation presentation. It is your responsibility to approach him / her.
Supervision

Your supervisor is your contact in all matters, both regarding the content, topics and the preparation presentation. It is your responsibility to approach him / her.

Informal meeting with the supervisor:

- ≥ 2 Weeks before the main presentation: discussion of the concepts to present
- ≥ 1 Week before the main presentation: discussion of the slides
- The latest till 15.6 (a month before the write-up submission): to discuss the content of the written document
- The latest till 19.8 (10 days before the final submission): Discussion of the corrected version of the document
Overview

1. Organisatorisches

- Ablauf
- Anforderungen

2. Topics

- Presentations
- Distribution
Topic Overview

1) Similarity of Notions
2) Syntax Trees
3) Text Matching
4) Text-variant graphs
5) Visualizations for Close Reading
6) Visualizations for Distant Reading
7) Fundamentals of Machine Learning and Topic Recognition
8) Topic Recognition via Latent Dirichlet Allocation
9) Topic Labeling using DBPedia
10) Text based Topic Labeling
1) Similarity of Notions

We all know which of these three objects are related...
1) Similarity of Notions

We all know which of these three objects are related...

But how the computer may learn it?
1) Similarity of Notions

We all know which of these three objects are related...

But how the computer may learn it?

It can ask google!
1) Similarity of Notions

We all know which of these three objects are related...

But how the computer may learn it?
It can ask google!

In the paper: Google Similarity Distance
  Theoretical background
  Experimental evaluation
2) Syntax Trees

Linguists studying natural language syntax, semantics and morphology describe their models using syntax trees.

- **S** - sentence
- **NP** - noun phrase
- **VP** - verb phrase
- **V** - verb
- **D** - articles
- **N** - noun

![Syntax Tree Example]

```
    S
    ├── NP
    │    ├── D
    │    │    └── N
    │    └── VP
    │         └── V
    └── N
```

Example: John hit the ball.
2) Syntax Trees

Linguists studying natural language syntax, semantics and morphology describe their models using syntax trees

- S - sentence
- NP - noun phrase
- VP - verb phrase
- V - verb
- D - articles
- N - noun

In the paper:

- An overview of algorithms for tree visualization
- Particular system TreeForm
- Evaluation
3) Text Matching

Researching texts, we often have to answer the question of how similar two pieces of text are.

The quick brown fox jumps over the lazy dog

Jump over the brown fox, lazy dog. Quick!
3) Text Matching

Researching texts, we often have to answer the question of how similar two pieces of text are.

The **quick** brown **fox** jumps over the **lazy** **dog**

Jump over the brown fox, lazy dog. Quick!
3) Text Matching

Researching texts, we often have to answer the question of how similar two pieces of text are.

The **quick** **brown fox** **jumps over the** **lazy dog**

Jump over the **brown fox**, **lazy dog**. **Quick!**
3) Text Matching

Researching texts, we often have to answer the question of how similar two pieces of text are.

The **quick** brown fox **jumps over the** lazy dog

Jump over the **brown fox**, lazy dog. **Quick!**
3) Text Matching

Researching texts, we often have to answer the question of how similar two pieces of text are.

The quick brown fox jumps over the lazy dog

Jump over the brown fox, lazy dog. Quick!

Block edit distance, the more similar are the texts the less the distance.
3) Text Matching

Researching texts, we often have to answer the question of how similar two pieces of text are.

The quick brown fox jumps over the lazy dog

Jump over the brown fox, lazy dog. Quick!

**Block edit distance**, the more similar are the texts the less the distance.

**In the paper:**
- Block-edit Distance
- Several models
- NP-completeness for some
- Algorithms for some
4) Text-Variant Graphs

In literary studies, researchers often need to compare several editions of the same text.
4) Text-Variant Graphs

In literary studies, researchers often need to compare several editions of the same text.

Text-variant graphs are the data structure that allows representation of several editions of the same text.
4) Text-Variant Graphs

In literary studies, researchers often need to compare several editions of the same text.

Text-variant graphs are the data structure that allows representation of several editions of the same text.

**In the paper:**
- Data-structure and its properties
- Operations on text-variant graphs
- Representation of text-variant graphs

Critical review of the algorithmic challenges. Overview of the follow up work.
Visualization for Close and Distant Reading

5) Close Reading


6) Distant Reading

openbible.info/blog/2009/03/phrase-net-bible-visualizations/
Visualization for Close and Distant Reading

5) Close Reading


6) Distant Reading


Rhymes | Character Clusters
---|---
Assonance | EH
Consonance | S
Vowel Slant Rhyme | EH1

NIGHT

Louise Bogan

The cold remote islands
And the blue situation
Where what breathes, breathes
The restless wind of the seafarers
And what drinks, drinks
The incoming tide:
Where shell and weed
Wait upon the salt wash of the sea
And the clear nights of stars
Swell their lights westward
To get behind the land;

openbible.info/blog/2009/03/phrase-net-bible-visualizations/
Visualization for Close and Distant Reading

5) Close Reading


6) Distant Reading


openbible.info/blog/2009/03/phrase-net-bible-visualizations/
Visualization for Close and Distant Reading

5) Close Reading

- "Night"
  Louise Bogan
  The cold remote islands
  And the blue situation
  Where what breathes, breathes
  The restless wind of the tides
  And what drinks, drinks
  The coming tide;
  Where shell and weed
  Wait upon the salt wash of the sea
  And the clear nights coil
  Swing their lights out
  To get behind the land.
  Coles, Meyer, Lein, McCurdy:
  "Empowering Play, Experimenting with Poems: Disciplinary Values and Visualization Development."

- "eMargin: A Collaborative Textual Annotation Tool. Ariadne 71, 2013"

6) Distant Reading

- "openbible.info/blog/2009/03/phrase-net-bible-visualizations/"


Visualization for Close and Distant Reading

5) Close Reading


6) Distant Reading

openbible.info/blog/2009/03/phrase-net-bible-visualizations/


Visualization for Close and Distant Reading

5) Close Reading

Paper:

On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges

Your Task:

- Give overview of presented techniques.
- Pick few and have a closer look:
  - Present research on algorithmic methods.
  - Identify open problems.


openbible.info/blog/2009/03/phrase-net-bible-visualizations/

6) Distant Reading
7) + 8) Automatic Topic Recognition

Task:

Topics of the text?

automatic extraction
7) + 8) Automatic Topic Recognition

Task:

Topics of the text?

automatic extraction

To solve this, it is helpful to first solve the inverse problem:

What do documents covering a given topic look like?
7) + 8) Automatic Topic Recognition

Task:

→ Need a (statistical) model of text covering a certain topic.

A particularly successful model: Latent Dirichlet Allocation

\((\approx 14000 \text{ citations})\)
7) + 8) Automatic Topic Recognition

Task:

Topics of the text?

automatic extraction

→ Need a (statistical) model of text covering a certain topic.
A particularly successful model: Latent Dirichlet Allocation

\( \approx 14000 \) citations

Your Task:

1. Fundamentals of Machine Learning and Topic Recognition
2. Topic Recognition via Latent Dirichlet Allocation
9) + 10) Topic Labelling

Documents

Labels
High Energy Physics
Physics
Quantum Mechanics
Particle Physics
9) + 10) Topic Labelling

Documents

Labels
High Energy Physics
Physics
Quantum Mechanics
Particle Physics

Topic Model
energy 0.2
atom 0.1
interaction 0.1
electron 0.04
quantum 0.02
...

?
9) + 10) Topic Labelling using DBPedia

Documents

Labels
High Energy Physics
Physics
Quantum Mechanics
Particle Physics

Concepts

Network Analysis

Topic Model
energy 0.2
atom 0.1
interaction 0.1
electron 0.04
quantum 0.02

. . .
9) + 10) Topic Labelling

using DBPedia

**Text-based**

**Documents**

**Candidate Labels**

**Labels**

- High Energy Physics
- Physics
- Quantum Mechanics
- Particle Physics

**Topic Model**

- energy: 0.2
- atom: 0.1
- interaction: 0.1
- electron: 0.04
- quantum: 0.02

**Concepts**

**Ranking/Matching**

**Network Analysis**

**DBpedia**
Topic Overview

1) Similarity of Notions
2) Syntax Trees
3) Text Matching
4) Text-variant graphs
5) Visualizations for Close Reading
6) Visualizations for Distant Reading
7) Fundamentals of Machine Learning and Topic Recognition
8) Topic Recognition via Latent Dirichlet Allocation
9) Topic Labeling using DBPedia
10) Text based Topic Labeling
Next Meetings

now:
    personal communication with your supervisor

12. May 9:45 am:
    Short Presentations

19. May 9:45 am:
    Presentation of two topics

2. June 9:45 am:
    Presentation of two topics
Next Meetings

now:
    personal communication with your supervisor

12. May 9:45 am:
    Short Presentations

19. May 9:45 am:
    Presentation of two topics

    always in this room

2. June 9:45 am:
    Presentation of two topics